UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/671,889	09/29/2003	Fred Gehrung Gustavson	YOR920030170US1	8009

48150   7590   11/02/2007

MCGINN INTELLECTUAL PROPERTY LAW GROUP, PLLC
8321 OLD COURTHOUSE ROAD
SUITE 200
VIENNA, VA 22182-3817

EXAMINER
VICARY, KEITH E

ART UNIT	PAPER NUMBER
2183	

MAIL DATE	DELIVERY MODE
11/02/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

	Application No.	Applicant(s)
Office Action Summary	10/671,889	GUSTAVSON ET AL.
	Examiner	Art Unit	
	Keith Vicary	2183	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE <u>3</u> MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.
- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
  Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on <u>03 October 2007</u>.

2a)☐ This action is **FINAL**.   2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) <u>*1-9 and 11-19*</u> is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) <u>*1-9 and 11-19*</u> is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☐ All   b)☐ Some * c)☐ None of:

      1.☐ Certified copies of the priority documents have been received.

      2.☐ Certified copies of the priority documents have been received in Application No. _____.

      3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1)☐ Notice of References Cited (PTO-892)

2)☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3)☒ Information Disclosure Statement(s) (PTO/SB/08)
    Paper No(s)/Mail Date <u>*7/31/2007*</u>.

4)☐ Interview Summary (PTO-413)
    Paper No(s)/Mail Date. _____.

5)☐ Notice of Informal Patent Application

6)☐ Other: _____.

DETAILED ACTION

1.   Claims 1-9 and 11-19 are pending in this office action and presented for

examination. Claims 1, 4, 6, 8, 11-12, 15, and 17-18 are newly amended and claims 10

and 20 are cancelled by amendment filed 9/14/2007.

Double Patenting

2.   Claims 1-9 and 11-19 of this application conflict with claims 1, 3-6, 8-12, and 14-

19 of Application No. 10671937. 37 CFR 1.78(b) provides that when two or more

applications filed by the same applicant contain conflicting claims, elimination of such

claims from all but one application may be required in the absence of good and

sufficient reason for their retention during pendency in more than one application.

Applicant is required to either cancel the conflicting claims from all but one application

or maintain a clear line of demarcation between the applications. See MPEP § 822.

3.   The nonstatutory double patenting rejection is based on a judicially created
doctrine grounded in public policy (a policy reflected in the statute) so as to prevent the
unjustified or improper timewise extension of the "right to exclude" granted by a patent
and to prevent possible harassment by multiple assignees.   A nonstatutory
obviousness-type double patenting rejection is appropriate where the conflicting claims
are not identical, but at least one examined application claim is not patentably distinct
from the reference claim(s) because the examined application claim is either anticipated
by, or would have been obvious over, the reference claim(s). See, e.g., *In re Berg*, 140
F.3d 1428, 46 USPQ2d 1226 (Fed. Cir. 1998); *In re Goodman*, 11 F.3d 1046, 29
USPQ2d 2010 (Fed. Cir. 1993); *In re Longi*, 759 F.2d 887, 225 USPQ 645 (Fed. Cir.
1985); *In re Van Ornum*, 686 F.2d 937, 214 USPQ 761 (CCPA 1982); *In re Vogel*, 422
F.2d 438, 164 USPQ 619 (CCPA 1970); and *In re Thorington*, 418 F.2d 528, 163
USPQ 644 (CCPA 1969).
        A timely filed terminal disclaimer in compliance with 37 CFR 1.321(c) or 1.321(d)
may be used to overcome an actual or provisional rejection based on a nonstatutory
double patenting ground provided the conflicting application or patent either is shown to

be commonly owned with this application, or claims an invention made as a result of activities undertaken within the scope of a joint research agreement.

Effective January 1, 1994, a registered attorney or agent of record may sign a terminal disclaimer. A terminal disclaimer signed by the assignee must fully comply with 37 CFR 3.73(b).

4.   Claims 1-9 and 11-19 are provisionally rejected on the ground of nonstatutory obviousness-type double patenting as being unpatentable over claims 1, 3-6, 8-12, and 14-19 of copending Application No. 10671937 in view of Gustavson et al. (Gustavson) (Superscalar GEMM-based Level 3 BLAS – The On-going Evolution of a Portable and High-Performance Library, Para'98, pages 207-215). Although the conflicting claims are not identical, they are not patentably distinct from each other because claims 1-9 and 11-19 of the instant application are obvious variants of claims 1, 3-6, 8-12, and 14-19 of the '937 application.

This is a provisional obviousness-type double patenting rejection.

5.   Claims 1-9 and 11-19 of the instant application contain every limitation of claims 1, 3-6, 8-12, and 14-19 of the '937 application; moreover, claims 1-9 and 11-19 of the instant application claim disclose inserting instructions to move data into said cache providing data into an FPU so that said LSUs can move said data into said Fregs in a timely manner for said linear algebra subroutine execution, whereas claims 1, 3-6, 8-12, and 14-19 of the '937 application merely claim preloading data into a floating point register of an FPU. Moreover, claims 1-9 and 11-19 of the instant application also disclose of data being prefetched into said cache from a memory in a nonstandard

format predetermined to reduce a number of data streams for a level 3 processing to be three streams and to allow a multiple loading of loads into said FPU by said LSU.

First, it would have been readily recognized by one of ordinary skill in the art at the time of the invention that the benefits of using cache in the instant application are numerous and include greater system performance due to the decreased access time to access cache in comparison to main memory combined with the locality of reference that is typical in most computer programs.

It would have been obvious to one of ordinary skill in the art at the time of the invention to implement cache into the instant application to gain greater system performance; it would have been readily recognized by one of ordinary skill in the art at the time of the invention that greater system performance is desirable in any processor. Furthermore, it would have been readily recognized by one of ordinary skill in the art at the time of the invention that this cache would fit into the '937 application by receiving data from the main memory and sending it to the floating point register, and that when preloading data into the floating point register in a system which uses a cache, that data would have to be prefetched into the cache in order to be preloaded into the register.

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to combine the widely-known teachings of cache with the invention of the '937 application in order to increase system performance.

Moreover; claims 1-9 and 11-19 of the instant application also disclose of data being prefetched into said cache from a memory in a nonstandard format predetermined to reduce a number of data streams for a level 3 processing to be three streams and to allow a multiple loading of loads into said FPU by said LSU.

On the other hand, Gustavson discloses, said data being prefetched into said cache from a memory in a nonstandard format (section 3.1, first indented paragraph of page 210, technique of keeping a small square block of C in registers; this technique of prefetching C in the format of a small square block as opposed to the prefetching of A and B can be considered nonstandard) to reduce a number of data streams for a level 3 processing to be three streams (section 3.1, first indented paragraph of page 210 as above, three total data streams are used, one for A, B, and C; note that as only a small square block of C instead of the entire C is being loaded into the registers, C is essentially a data stream of small square blocks. Also note that streams can be broadly read to be the data from the FPU registers to the FPU itself and thus encompasses A, B, and C regardless of the above technique) and to allow a multiple loading of loads into said FPU by said LSU (section 3.1, first indented paragraph of page 210 as above, number of load and store instructions; there thus must exist multiple loads into said FPU by said LSU).

Gustavson's teaching above maximizes the ratio between the number of MAAs and the number of load and store instructions used to transfer data to and from registers (section 3.1, page 210, first indented paragraph, first 5 lines).

It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the teaching of Gustavson with the invention of the '937 application in order to maximize the ratio between the number of MAAs and the number of load and store instructions, which enables the increase in system performance. It would have been readily recognized to one of ordinary skill in the art at the time of the invention that the teaching of Gustavson does not render the invention of the '937

application unusuable. The claims of the '937 application disclose of preloading data to

an FPU for linear algebra operations so that the data may be timely executed by the

FPU but does not disclose the format of the data or the format of how the preloading is

actually done. Gustavson teaches the above limitations in describing how to gain an

increase in system performance when executing linear algebra operations.

Therefore, it would have been obvious to one of ordinary skill in the art at the

time of the invention to combine the teaching of Gustavson with the invention of the '937

application in order to maximize the ratio between the number of MAAs and the number

of load and store instructions, which enables the increase in system performance

a.   Further note that claims 2, 11, and 13 in the instant application also claim

that prefetching data is accomplished by utilizing time slots caused by a

difference between a time to execute instructions in said subroutine execution

process and a time to load said data, while claims 1, 11, and 12 of the '937

application does not explicitly disclose this.

It would have been readily recognized by one of ordinary skill in the art at

the time of the invention that prefetching data in general cuts down the amount of

time a processor is waiting for a memory miss to be serviced, and prefetching by

utilizing time slots caused by a difference between a time to execute instructions

and a time to load said data allows for data to be prefetched ahead of time

without delaying any other instructions that are being processed. Furthermore, it

would have been readily recognized by one of ordinary skill in the art at the time

of the invention that the benefits of prefetching are contingent upon other

instructions not being delayed due to the prefetching; thus, it would have been

readily recognized to one of ordinary skill in the art at the time of the invention

that prefetching would be done by utilizing these time slots of inactivity.

Therefore, it would have been obvious to one of ordinary skill in the art at

the time of the invention to combine the widely-known method of prefetching by

utilizing time slots with the '937 application in order to cut down the amount of

time a processor is waiting for a memory miss to be serviced, thus increasing

overall system performance.

6.   Aside from the obvious variants listed above, claim 1 of the '937 application

contains every element of claim 1 of the instant application.

7.   Aside from the obvious variants listed above, claim 1 of the '937 application

contains every element of claim 2 of the instant application.

8.   Aside from the obvious variants listed above, claim 3 of the '937 application

contains every element of claim 3 of the instant application.

9.   Aside from the obvious variants listed above, claim 4 of the '937 application

contains every element of claim 4 of the instant application.

10.   Aside from the obvious variants listed above, claim 5 of the '937 application

contains every element of claim 5 of the instant application.

11.   Aside from the obvious variants listed above, claim 6 of the '937 application

contains every element of claim 6 of the instant application.

12.   Aside from the obvious variants listed above, claim 8 of the '937 application

contains every element of claim 7 of the instant application.

13.   Aside from the obvious variants listed above, claim 9 of the '937 application

contains every element of claim 8 of the instant application.

14.   Aside from the obvious variants listed above, claim 10 of the '937 application contains every element of claim 9 of the instant application.

15.   Aside from the obvious variants listed above, claim 6 of the '937 application contains every element of claim 11 of the instant application.

16.   Aside from the obvious variants listed above, claim 12 of the '937 application contains every element of claim 12 of the instant application.

17.   Aside from the obvious variants listed above, claim 12 of the '937 application contains every element of claim 13 of the instant application.

18.   Aside from the obvious variants listed above, claim 14 of the '937 application contains every element of claim 14 of the instant application.

19.   Aside from the obvious variants listed above, claim 15 of the '937 application contains every element of claim 15 of the instant application.

20.   Aside from the obvious variants listed above, claim 16 of the '937 application contains every element of claim 16 of the instant application.

21.   Aside from the obvious variants listed above, claim 17 of the '937 application contains every element of claim 17 of the instant application.

22.   Aside from the obvious variants listed above, claim 18 of the '937 application contains every element of claim 18 of the instant application.

23.   Aside from the obvious variants listed above, claim 19 of the '937 application contains every element of claim 19 of the instant application.

Claim Rejections - 35 USC § 112

24.   The following is a quotation of the first paragraph of 35 U.S.C. 112:

> The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

25.   The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

26.   Claims 1-9 and 11-19 are rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention.

27.   Claims 1, 6, 12, and 17 as amended recites the limitation "data...in a nonstandard format" in line 9. The "nonstandard" limitation does not appear to be present in the original instant application. If the limitation is present somewhere in one of the co-pending applications, this should be noted in any subsequent arguments to overcome the rejection.

28.   Claims 1, 6, 12, and 17 as amended recites the limitation "a nonstandard format predetermined to reduce a number of data streams for a level 3 processing to be three streams" in lines 8-10. This limitation does not appear to be present in the original instant application. If the limitation is present somewhere in one of the co-pending applications, this should be noted in any subsequent arguments to overcome the rejection.

29.   Claims 1, 6, 12, and 17 as amended recites the limitation "allow a multiple loading of loads" in the second to last line. The limitation, which encompasses the broad interpretation of loading multiple registers with one instruction, does not appear to

be present in the original instant application. If the limitation is present somewhere in

one of the co-pending applications, this should be noted in any subsequent arguments

to overcome the rejection.

      b.   Claims 2-5, 7-9, 11, 13-16, and 18-19 are rejected for failing to alleviate

the rejection of claims 1, 6, 12, and 17 above.

30.   Claims 2, 11, and 13 as amended recites the limitation "existing in a Level 3

Dense Linear Algebra Subroutine" or a variation thereof. The limitation does not appear

to be present and connected to the other limitations in the instant or incorporated

specifications and thus is considered new matter. If this Level 3 Dense Linear Algebra

Subroutine is a subset of Level 3 BLAS, then the scope of the claim is being narrowed

with no basis in the specification. If this Level 3 Dense Linear Algebra Subroutine is

synonymous with Level 3 BLAS or something else in the specifications, then a

reference or citation should be provided which validates this.

31.   Claims 1-20 are rejected under 35 U.S.C. 112, second paragraph, as being

indefinite for failing to particularly point out and distinctly claim the subject matter which

applicant regards as the invention.

32.   Claims 1, 6, 12, and 17 recite the limitation "timely moved" or "timely manner." It

is indefinite as to what this limitation implies. Although timely movement in the context

of the claim can be logically construed to be movement before the data is needed (due

to the prefetching limitation), it is indefinite as to whether "timely" movement is also

being used to mean that the movement is, for example, done right before the data is

needed, or whether "timely" movement is also being used to mean that the movement is done as soon in advance as possible. If "timely" movement does not cover either of the aforementioned examples and is only used to describe general prefetching, it is unclear as to what the purpose of the limitation is as it appears to be redundant.

33.   Claims 1, 6, 12, and 17 recite the limitation "in a nonstandard format" in, for example, line 9 of claim 1. It is indefinite as to what exactly makes a format "nonstandard".

34.   Claims 1, 6, 12, and 17 recite the limitation "level 3 processing" in, for example, line 10 of claim 1. It is indefinite as to what exactly a "level 3 processing" is.

35.   Claims 1, 6, 12, and 17 recite the limitation "multiple loading of loads" in, for example, lines 10-11 of claim 1. It is indefinite as to whether "loads" is being used to refer to load instructions/micro-ops or data.

     c.   Claims 2-5, 7-9, 11, 13-16, and 18-19 are rejected for failing to alleviate the rejection of claims 1, 6, 12, and 17 above.

36.   Claim 6 recites the limitation "wherein said matrix data in said memory is timely moved by inserting moving instructions to be loaded into said cache" in lines 7-8. The limitation as written seems to imply that it is the moving instructions which are being loaded into said cache and not said matrix data, and should thus be rewritten to be more clear.

     d.   Claims 7-9 and 11 are rejected for failing to alleviate the rejection of claim 6 above.

Claim Rejections - 35 USC § 102

1.   The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that

form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

2.   Claims 1-9 and 11-19 are rejected under 35 U.S.C. 102(b) as being anticipated

by Gustavson et al. (Gustavson) (Superscalar GEMM-based Level 3 BLAS – The On-

going Evolution of a Portable and High-Performance Library, Para'98, pages 207-215).

3.   Consider claims 1 and 12, Gustavson discloses for an execution code (section 1,

line 6, BLAS code) controlling an operation of said floating point unit (FPU) (section 3.1,

line 4, discloses floating point registers, therefore it is inherent there are floating point

units that are doing the multiplications as in section 1, line 2) performing a linear algebra

subroutine execution (section 1, line 8, routine along with section 1, line 1, linear

algebra), inserting instructions to move data into said cache providing data for said FPU

so that said LSUs can move said data into said Fregs in a timely manner for said linear

algebra subroutine execution (section 4.1, line 8, algorithmic prefetching), said data

being prefetched into said cache from a memory in a nonstandard format (section 3.1,

first indented paragraph of page 210, technique of keeping a small square block of C in

registers; this technique of prefetching C in the format of a small square block as

opposed to the prefetching of A and B can be considered nonstandard) to reduce a

number of data streams for a level 3 processing to be three streams (section 3.1, first

indented paragraph of page 210 as above, three total data streams are used, one for A,

B, and C; note that as only a small square block of C instead of the entire C is being

loaded into the registers, C is essentially a data stream of small square blocks. Also

note that streams can be broadly read to be the data from the FPU registers to the FPU

itself and thus encompasses A, B, and C regardless of the above technique) and to

allow a multiple loading of loads into said FPU by said LSU (section 3.1, first indented

paragraph of page 210 as above, number of load and store instructions; there thus must

exist multiple loads into said FPU by said LSU. Also see the second-to-last paragraph

of section 3.1, multiple element load instructions).

4.   Consider claim 6, Gustavson discloses an apparatus, comprising: a memory to

store matrix data to be used for processing in a linear algebra program (section 4, line

12, shared main memory and section 4.2, lines 7-9, elements of the matrix); a floating

point unit (FPU) to perform said processing (section 3.1, line 4, discloses floating point

registers, therefore it is inherent there are floating point units that are doing the

multiplications as in section 1, line 2); a load/store unit (LSU) to load data to be

processed by said FPU (section 3.1, lines 6-7, load and store operations, thus it is

inherent there is a load/store unit), said LSU loading said data into a plurality of floating

point registers (FRegs) (section 3.1, line 4, floating point registers); and a cache to store

data from said memory and provide said data to said Fregs (section 4.1, line 4, cache),

wherein said matrix data in said memory is timely moved by inserting moving

instructions to be loaded into said cache prior to a need for said data to be loaded by

said LSU into said Fregs for said processing, (section 4.1, line 8, algorithmic

prefetching), said data being prefetched into said cache from a memory in a

nonstandard format (section 3.1, first indented paragraph of page 210, technique of

keeping a small square block of C in registers; this technique of prefetching C in the

format of a small square block as opposed to the prefetching of A and B can be

considered nonstandard) predetermined to reduce a number of data streams for a level

3 processing to be three streams (section 3.1, first indented paragraph of page 210 as

above, three total data streams are used, one for A, B, and C; note that as only a small

square block of C instead of the entire C is being loaded into the registers, C is

essentially a data stream of small square blocks. Also note that streams can be broadly

read to be the data from the FPU registers to the FPU itself and thus encompasses A,

B, and C regardless of the above technique) and to allow a multiple loading of loads into

said FPU by said LSU (section 3.1, first indented paragraph of page 210 as above,

number of load and store instructions; there thus must exist multiple loads into said FPU

by said LSU. Also see the second-to-last paragraph of section 3.1, multiple element

load instructions).

5.   Consider claim 17, Gustavson discloses a method of providing a service

involving at least one of solving and applying a scientific/engineering problem, said

method comprising at least one of:

using a linear algebra software package that computes one or more matrix

subroutines, wherein said linear algebra software package generates an execution code

(section 1, line 6, BLAS code) controlling an operation of a floating point unit (FPU)

(section 3.1, line 4, discloses floating point registers, therefore it is inherent there are

floating point units that are doing the multiplications as in section 1, line 2) performing a

linear algebra subroutine execution (section 1, line 8, routine along with section 1, line

1, linear algebra), said data being prefetched into said cache from a memory in a nonstandard format (section 3.1, first indented paragraph of page 210, technique of keeping a small square block of C in registers; this technique of prefetching C in the format of a small square block as opposed to the prefetching of A and B can be considered nonstandard) to reduce a number of data streams for a level 3 processing to be three streams (section 3.1, first indented paragraph of page 210 as above, three total data streams are used, one for A, B, and C; note that as only a small square block of C instead of the entire C is being loaded into the registers, C is essentially a data stream of small square blocks. Also note that streams can be broadly read to be the data from the FPU registers to the FPU itself and thus encompasses A, B, and C regardless of the above technique) and to allow a multiple loading of loads into said FPU by said LSU (section 3.1, first indented paragraph of page 210 as above, number of load and store instructions; there thus must exist multiple loads into said FPU by said LSU).

providing a consultation for solving a scientific/engineering problem using said linear algebra software package (it is inherent that the BLAS will solve some type of scientific/engineering problem for someone who may or may not be the operator of the BLAS program); transmitting a result of said linear algebra software package on at least one of a network, a signal-bearing medium containing machine-readable data representing said result, and a printed version representing said result; and receiving a result of said linear algebra software package on at least one of a network, a signal-bearing medium containing machine-readable data representing said result, and a printed version representing said result (it is inherent that the result of the problem will be conveyed to someone who may or may not be the operator of the BLAS program;

furthermore, it is inherent that the result can only be shown either through a printout or through some type of electronic means, which encompasses voice through a phone or data through a network that is read via a monitor).

6.   Consider claims 2, 11, and 13, Gustavson discloses said timely moving data is accomplished by scheduling move type instructions into time slots existing in a Level 3 Dense Linear Algebra Subroutine. As explained above, it is inherent to prefetching that data is loaded into the cache before the instruction that needs that data is executed, thus there must be a difference between the time of that instruction execution and the time of its data loading, otherwise it would not be prefetching. Furthermore, Gustavson discloses in page 12, lines 2-3 of section 4.1 that the prefetching instruction does not disturb ongoing computations and data references, thus this prefetching must be done in "time slots" which are independent of other instruction fetching.   Gustavson in section 3, line 5, discloses of DGEMM, which is a type of Level 3 Dense Linear Algebra Subroutine.

7.   Consider claims 3, 7, and 14, Gustavson discloses said linear algebra subroutine comprises a matrix multiplication operation (section 1, line 2, matrix multiply).

8.   Consider claims 4, 8, 15, and 18, Gustavson discloses said matrix subroutine comprises an equivalent of a subroutine from a LAPACK (Linear Algebra PACKage) (section 1, line 1, discloses a BLAS, which is a part of LAPACK).

9.   Consider claims 5, 9, 16, and 19, Gustavson discloses said linear algebra

subroutine comprises a BLAS Level 3 L1 cache kernel (Abstract, lines 1-6, level 3 BLAS

kernel and level 1 cache).

Response to Arguments

37.   Examiner rescinds his written description rejection related to the limitation "timely

moving" due to the newly amended limitations which tie "timely moving" to prefetching.

Because timely moving must fall under the scope of prefetching, due to the way the

limitations are written in the claim, the limitation "timely moving" no longer appears to be

new matter. However, a 112 indefiniteness rejection does arise in using both

limitations; see the 112 rejection section above.

38.   Applicant has cited support in page 10 of the arguments for the non-standard

format and three data streams as being found in lines 21 of page 14 through lines 12 of

page 15 of the instant specification. However, this citation seems to imply that there are

two streaming matrices and not three, and examiner cannot find any indication that a

"nonstandard" format is used, as it is also indefinite as to what constitutes

nonstandardness.

39.   Applicant argues on page 10 that support for the terminology "Level 3 Dense

Linear Algebra Subroutine" is found in various locations in the original specification

because DGEMM is one example of a level 3 dense linear algebra subroutine.

However, the limitation is nevertheless new matter. The original specification covers

DGEMM in particular but does not disclose that the invention may be applicable to *any*

Level 3 Dense Linear Algebra Subroutine. Even though it may be apparent that this is

possible, it is nevertheless considered new matter. Page 13, lines 10-15, discloses that

the approach which is presented as it relates to the DGEMM may also be extended to

the other Level 3 BLAS and matrix operation routine, but as "Level 3 Dense Linear

Algebra Subroutines" are not explicitly disclosed in the specification, it is considered

new matter.

40.   Applicant argues on page 11 that applicants believe that incorporation into the

independent claims of the description of the newer computer architectures distinguishes

from the discussion in this newly-cited reference. However, examiner is uncertain as to

what newly amended limitations encompass this description of the newer computer

architectures.

41.   Examiner has re-read interview slides and co-pending applications which seem

to imply that the novelty of the overall method of the instant specifications is related to

adapting linear algebra operations to newer processor architectures, e.g., processor

architectures which *already* support multiple loads or quad loads, and prefetching. In

other words, due to *existing* processing architectures, "non-standard" data structures

and register blocking methods are used for implementing linear algebra operations.

With this analysis, and noting that the newly amended limitations seem directed toward

linear algebra data structures (nonstandard format) and methods for using an existing

processor (reduce a number of data streams for a level 3 processing), examiner advises

applicant to amend in limitations in any further prosecution that specifically relate to any

improvements in processor architecture itself, as it appears that otherwise, additional

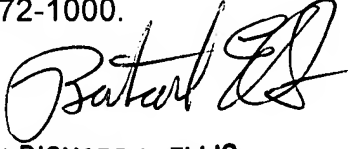double patenting will occur with co-pending applications that disclose the matrix data

format and so forth in view of any art which discloses the well-known concept of pre-

fetching.

Conclusion

42.   Any inquiry concerning this communication or earlier communications from the

examiner should be directed to Keith Vicary whose telephone number is (571) 270-

1314. The examiner can normally be reached on Monday - Friday, 8:00 a.m. - 5:00

p.m., EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's

supervisor, Eddie Chan can be reached on 571-272-4162. The fax phone number for

the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the

Patent Application Information Retrieval (PAIR) system. Status information for

published applications may be obtained from either Private PAIR or Public PAIR.

Status information for unpublished applications is available through Private PAIR only.

For more information about the PAIR system, see http://pair-direct.uspto.gov. Should

you have questions on access to the Private PAIR system, contact the Electronic

Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a

USPTO Customer Service Representative or access to the automated information

system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

RICHARD L. ELLIS
PRIMARY EXAMINER

kv